

Weekly Report

Junhua Lu

2016 年 9 月 25 日

Done

- Many minor changes on the vis paper and figures.
- Read several parts of a survey on unsupervised rare pattern mining (association rule).
- Revised our second patent and handed it to the lawyer.

To do

- Finish reading the survey.
- Think of new possible topic.

论文

Unsupervised Rare Pattern Learning: A survey 这篇文章主要介绍的是rare pattern mining, 里面介绍的方法都属于关联规则一类的. 看这篇文章一个目的是因为我们后面准备做游戏数据异常发现, 诸如作弊之类. 但看了文章一部分后我也思考了关联规则和一般的数据挖掘方法(anomaly detection)以及机器学习之类确实不太一样. 挖掘类的一般都是一个数据点有着相同的维度但是维度上数据不同, 在这里挖掘; 关联规则则是类别型的不定维的(当然要强行转换过去成数据点也不是不可以); 挖掘可以分为有监督半监督无监督的, 但是关联规则一般是无监督. 从输出上来说, 一般挖掘的输出是label或者score, 而关联规则输出则是规则.

在静态数据的static rare pattern mining中, 分为可变support阈值, 无support阈值, 结果限制, 和基于树的方法. 关联规则中, 往往有一个固定的support的要求, 这常常把一些有用的规则一刀切切走了, 或者把一些rare pattern给淹没了. 可变阈值就是用各种技巧让不同item的阈值变得不一样; 无support阈值则多是采用一个相似性的指标或者基于confident的找规则的方法; 结果限制则是比较功利的找出必须包含某结果的规则. 后面部分尚未研读, 后面会解决.

此外我向刘柏了解, 他现在也在某一个浙大老师的作弊检测项目中, 他有已经标记好作弊的数据(其余数据据悉和我们的类似). 具体能否给我们尚不清楚, 后面还要进一步商量或者在我们定好后面的题目以后请老师们去交涉一下.